

ABSTRACT

There is presently no single accepted scheme to support Quality of Service guarantees for IP based applications, although RSVP is the prevailing standard for network resource reservation. Thus, providing end-to-end Quality of Service guarantees for content delivery across the Internet presents a challenge for Service Providers. Typically load balancing algorithms are used for picking the best server, with respect to server capacity for satisfying a request, but they can not guarantee Quality of Service for the whole duration of the network transaction. The invention proposes a method of distributing server load in an IP network, comprising building an association between a subscriber edge device and a server controller using a packet switched network Quality of Service mechanism. A fair share of server bandwidth is offered out to the subscriber edge device, and a resource request from the client is directed, via the subscriber edge device and through a server controller, to a server having an amount of server bandwidth required by the client. The required fair share of server bandwidth is then reserved for meeting the client's resource request. The invention also proposes a subscriber edge device, a data center device, and a communications network comprising each. The subscriber edge device comprises, a resource requester for sending a request to the server controller associated with the source location of the requested resource, a resource reserver that reserves an amount of bandwidth using a Quality of Service mechanism and releases any unneeded bandwidth, and a resource returner that returns the requested resource to the client. The data center device comprises, a resource allocator that allocates fair shares of server bandwidth to the network, and a server controller that offers fair shares of server bandwidth using a Quality of Service mechanism.